ERDC/CRREL TR-19-13

Cold Regions Research and Engineering Laboratory

US Army Corps of Engineers®
Engineer Research and Development Center

ERDC
ENGINEER RESEARCH & DEVELOPMENT CENTER

Pavement-Transportation Computer Assisted Structural Engineering (PCASE)

WorldIndex Database Update 2018

Lynette A. Barna, Steven F. Daly, Kathleen F. Jones, and John J. Gagnon

July 2019

The U.S. Army Engineer Research and Development Center (ERDC) solves the nation's toughest engineering and environmental challenges. ERDC develops innovative solutions in civil and military engineering, geospatial sciences, water resources, and environmental sciences for the Army, the Department of Defense, civilian agencies, and our nation's public good. Find out more at www.erdc.usace.army.mil.

To search for other technical reports published by ERDC, visit the ERDC online library at http://acwc.sdp.sirsi.net/client/default.

ERDC/CRREL TR-19-13
July 2019

WorldIndex Database Update 2018

Lynette A. Barna, Steven F. Daly, Kathleen F. Jones, and John J. Gagnon

U.S. Army Engineer Research and Development Center (ERDC)
Cold Regions Research and Engineering Laboratory (CRREL)
72 Lyme Road
Hanover, NH 03755-1290

Final Report

Prepared for U.S. Army Corps of Engineers, Transportation Systems Center
1616 Capitol Avenue
Omaha, NE 68102-4901

Under FAN S0140441

Abstract

The primary software program for designing and evaluating pavements on military installations is the Pavement-Transportation Computer Assisted Structural Engineering (PCASE) program. PCASE, which is undergoing an update to version PCASE7, is an engineering tool for the design and evaluation of airfields and roadways using Department of Defense criteria. The WorldIndex database is a necessary component of PCASE and is integrated into it. The WorldIndex database internally supplies PCASE with the climate parameters required to determine the maximum depth of frost penetration. The database contains other temperature-based parameters used in structural evaluations.

This report documents a recent update to the WorldIndex database that uses historical surface air-temperature observations from 1980 to 2017 at over 10,000 locations around the globe. The database contains 80 air-temperature-based parameters determined for each station, including the three parameters that PCASE requires to compute the maximum depth of frost penetration for pavement design in cold regions: the average annual air temperature, the average freezing degree-days, and the mean length of the freezing season. The report concludes with recommendations for future updates.

Contents

Figures and Tables

Figures

Tables

Preface

This study was conducted for the U.S. Army Corps of Engineers (USACE) Transportation Systems Center under FAN S0140441. The technical monitors were Mr. George VanSteenburg, USACE Transportation Systems Center, and Dr. Carlos Gonzalez, U.S. Army Engineer Research and Development Center (ERDC), Geotechnical and Structures Laboratory (CEERD-GMA).

The work was performed by the Force Projection and Sustainment Branch (CEERD-RRH), the Terrestrial and Cryospheric Sciences Branch (CEERD-RRG), and the Engineering Resources Branch (CEERD-RRE) of the Research and Engineering Division (CEERD-RR) and the Terrain and Ice Engineering Group of the Remote Sensing / Geographic Information Systems Center of Expertise (CEERD-RS), ERDC Cold Regions Research and Engineering Laboratory (ERDC-CRREL). At the time of publication, Dr. Justin Putnam was Acting Chief, CEERD-RRH; Dr. John Weatherly was Chief, CEERD-RRG; Dr. Caitlin Callaghan was Acting Chief, CEERD-RRE; Mr. Jared Oren was Acting Chief CEERD-RR; Mr. Stephen Newman was lead, Terrain and Ice Engineering Group; Mr. David Finnegan was Chief, CEERD-RS; and Dr. Robert E. Davis, CEERD-RZT, was the Technical Director. The Deputy Director of ERDC-CRREL was Mr. David B. Ringelberg, and the Director was Dr. Joseph L. Corriveau.

The authors gratefully acknowledge technical review by Dr. John Weatherly of CRREL and Dr. Edel Cortez, retired CRREL.

COL Ivan P. Beckman was the Commander of ERDC, and Dr. David W. Pittman was the Director.

Acronyms and Abbreviations

CDD	Cumulative Degree-Days
CI	Condition Index
COOP	Cooperative Observer Program
CRREL	Cold Regions Research and Engineering Laboratory
DFI	Design Freezing Index
ERDC	U.S. Army Engineer Research and Development Center
FDD	Freezing Degree-Days
GHCN	Global Historical Climatology Network
GSOD	Global Surface Summary of Day
HEC-DSS	Hydrologic Engineering Center Data Storage System
IR	Internal Review
ISD	Integrated Surface Database
MEPDG	Mechanistic-Empirical Pavement Design Guide
MERRA-2	Modern Era Retrospective analysis for Research and Applications, Version 2
NASA	National Aeronautics and Space Administration
NCEI	National Centers for Environmental Information
NOAA	National Oceanic and Atmospheric Administration
PCASE	Pavement-Transportation Computer Assisted Structural Engineering
PR	Programming
SD	Station Description
UFC	Unified Field Criteria

USACE	U.S. Army Corps of Engineers
USAF/WMO	U.S. Air Force / World Meteorological Organization
WBAN	Weather Bureau Army Navy

1 Introduction

1.1 Background

Pavement-Transportation Computer Assisted Structural Engineering (PCASE) is software for designing and evaluating airfields and roadways according to Tri-Service Criteria (Adolf 2010). It is the primary tool used for these applications at military installations and is mandatory when designing roads and parking areas trafficked by special military vehicles. Outside of the United States and its territories and possessions, its use is mandatory when designing roads and parking areas on U.S. military installations for all vehicle types (Naval Facilities Engineering Command 2016).

The WorldIndex database provides global climate data that PCASE requires and is integrated into the PCASE software. For example, PCASE uses the modified Berggren solution (Cortez et al. 2000) to estimate the maximum frost depth at locations that experience seasonal freezing and thawing. The PCASE user interface is used to select a location, and then PCASE interrogates the WorldIndex database to obtain the air-temperature data required to calculate the maximum frost depth. The WorldIndex database has 80 air-temperature-based parameters, including those used to calculate the maximum frost depth.

The previous version of PCASE, PCASE2.09, was recently upgraded. The new version is referred to as PCASE7. As part of the upgrade to PCASE7, the U.S. Army Cold Regions Research and Engineering Laboratory (CRREL) updated the WorldIndex database to reflect more recent ground-based air-temperature observations. It has been about 20 years since the last update to the WorldIndex database. The previous update included nearly 30 years of observations, ending in 1995 (WorldIndex 1999). The new update is based on weather data from 1980 to 2017 at over 16,000 weather stations around the world. This report documents the update of the WorldIndex database for the PCASE 7 upgrade.

1.2 Objectives

The primary objective of this study was to update the WorldIndex database using recent surface observations made at all available stations across the globe. To achieve this goal, CRREL used consistent, documented methodologies to determine the database values; preserved the database schema of the previous version of the WorldIndex database; and developed procedures that could serve as a basis for automating future database updates.

1.3 Approach

This report begins with a description of the WorldIndex database. The description defines each parameter included in the database and provides example values. Sections 3 and 4 describe the two data sources used in the update and the framework of the data processing. Section 5 discusses the continuity and consistency of stations from the previous WorldIndex database to the updated WorldIndex database. Finally, section 6 provides the algorithms used to estimate each WorldIndex temperature parameter.

2 WorldIndex Database

The WorldIndex database can be thought of as a large table with columns and rows. Each column is either station metadata or an air-temperature parameter. Each row has values for a particular station. The header of each column identifies the parameter in that column. There are four types of column headers (defined in Table 1). There are 99 total headers in the updated WorldIndex database (listed and defined in Table 2). In Table 2, column 1 is the collection of letters used as the WorldIndex parameter identifier that is used as a column header. The WorldIndex parameter identifier is a condensed version of the parameter definition. For consistency, it is expressed here in all capital letters. The parameter identifier is the same as the identifier name in the previous version of the database, if the parameter was included in the previous version. Column 2 defines the parameter, including units and format of values. The values of the parameters are whole numbers if a format is not specified. Column 3 is the parameter type as defined in Table 1. Column 4 shows example data based on Minot, North Dakota. The units of all numeric values are metric.

Table 1. WorldIndex column types definitions.

Acronym	Name	Definitions
SD	Station description	These columns provide data about the station being analyzed. This information is directly from National Centers for Environmental Information.
PCASE	Required for PCASE database	These columns are the results of the analysis of the air-temperature data provided by the ground observation stations.
IR	Internal review	These columns provide data summary information to determine data accuracy.
PR	Programming	These columns are needed for the software programming structure.

Table 2. WorldIndex column header definitions.

WorldIndex Parameter Identifier Used as Column Header	Definition	Type	Minot, ND, Value
SYS_DELETED	Station data in previous database versions.	PR	FALSE
UID_GHCN	Unique station identification number.	PR	USW0002413
ID	Row number in the MS Access database. Unitless.	SD	<blank>

WorldIndex Parameter Identifier Used as Column Header	Definition	Type	Minot, ND, Value
SOURCE	Legacy identifier listed as G, T, or <blank>. Unitless.	SD	<blank>
WMO_NUMBER	Legacy international station identifier from the World Meteorological Organization. Unitless.	SD	727676
FILENAME	Legacy combination 8-character alphanumeric name in an MS Excel file format. The field begins with the characters FC or FD, followed by the station number, a "G" or "T" designator (only used for international stations), and .xls. Unitless.	SD	<blank>
STN_ID	Station identification number. Unitless.	SD	24013
STN_NAME	Station identification name. Unitless.	SD	MINOT_INTERNATIONAL_ARPT
COUNTRY	For U.S. stations, the state where the station is located; for international stations, the country where the station is located. Unitless.	SD	North Dakota
ELEVATION	Vertical height of the station above sea level. Meters.	SD	507.5
YRS_TAVG	Number of years of data used to calculate the average air temperature. Years.	PCASE	37
YRS_FDD	Number of years of data used to calculate the freezing degree-days. Years.	PCASE	37
FRACLAT	Station geographic latitude coordinate. Decimal format. Degrees.	SD	48.255
FRACLON	Station geographic longitudinal coordinate. Decimal format. Degrees.	SD	−101.273
YRLY_TAVG	Average of the annual average temperatures over the period of record. Degrees Celsius. Tenths.	PCASE	6.4
YRLY_STDEV_TAVG	Standard deviation of the annual average temperatures. Degrees Celsius. Tenths.	PCASE	1.3

WorldIndex Parameter Identifier Used as Column Header	Definition	Type	Minot, ND, Value
YRLY_AVG_FDD	Average annual maximum cumulative freezing degree-days. Celsius degree-days. Tenths.	PCASE	989.7
YRLY_STDEV_FDD	Standard deviation of annual maximum cumulative freezing degree-days. Celsius degree-days. Tenths.	PCASE	309.6
FDD_PCT95	95th percentile of the annual maximum freezing degree-days. Celsius degree-days. Tenths.	PCASE	1472.6
FDD_PCT90	90th percentile of the annual maximum freezing degree-days. Celsius degree-days. Tenths.	PCASE	1452
FDD_PCT50	50th percentile of the annual maximum freezing degree-days. Celsius degree-days. Tenths.	PCASE	1044
FDD_PCT10	10th percentile of the annual maximum freezing degree-days. Celsius degree-days. Tenths.	PCASE	523.3
FDD_PCT05	5th percentile of the annual maximum freezing degree-days. Celsius degree-days- Tenths.	PCASE	462.4
FDD_PCT03	3rd percentile of the annual maximum freezing degree-days. Celsius degree-days. Tenths.	PCASE	462.4
JAN_TAVG_AVERAGE	Average of the average daily air temperature for January (as January represents the coldest month of the year for the northern hemisphere). Degrees Celsius. Tenths.	PCASE	−10.1
JAN_TMIN_AVERAGE	Average of the minimum daily temperature for January (as January represents the coldest month of the year for the northern hemisphere). Degrees Celsius. Tenths.	PCASE	−15.0
JAN_TMAX_AVERAGE	Average of the maximum daily temperature for January (as January represents the coldest month of the year for the northern hemisphere). Degrees Celsius. Tenths.	PCASE	−5.3
JUL_TAVG_AVERAGE	Average of the average daily temperature for July (as July represents the warmest month of the year for the northern hemisphere). Degrees Celsius. Tenths.	PCASE	21.7

WorldIndex Parameter Identifier Used as Column Header	Definition	Type	Minot, ND, Value
JUL_TMIN_AVERAGE	Average of the minimum daily temperature for July (as July represents the warmest month of the year for the northern hemisphere). Degrees Celsius. Tenths.	PCASE	14.8
JUL_TMAX_AVERAGE	Average of the maximum daily temperature for July (as July represents the warmest month of the year for the northern hemisphere). Degrees Celsius. Tenths.	PCASE	28.5
HI_TMAX	Maximum daily maximum air temperature during the period of record. Degrees Celsius. Tenths.	PCASE	40.6
LO_TMIN	Minimum daily minimum air temperature during the period of record. Degrees Celsius. Tenths.	PCASE	−37.8
YRS_FREEZE	Number of years in the period of record used to determine the freezing-season parameters. Freezing season is the portion of the year when the daily average air temperature is below the freezing point of water. Year.	PCASE	37
AVERAGE_START	Average start day of the freezing season counted from 1 August for locations in the northern hemisphere or from1 February for locations in the southern hemisphere. Start date is when the daily average air temperature falls below the freezing point of water. Day. Whole number.	PCASE	103
STDEV_START	Standard deviation of the freezing-season start day. Day. Tenths.	PCASE	12.0
LATE_START	Latest start day of the freezing season determined over the period of record counted from 1 August for locations in the northern hemisphere or from 1 February for locations in the southern hemisphere. Day. Whole number.	PCASE	133
EARLY_START	Earliest start day of the freezing season determined over the period of record counted from 1 August for locations in the northern hemisphere or from 1 February for locations in the southern hemisphere. Day. Whole number.	PCASE	80

WorldIndex Parameter Identifier Used as Column Header	Definition	Type	Minot, ND, Value
AVERAGE_STOP	Average stop day of the freezing season determined over the period of record counted from 1 August for locations in the northern hemisphere or from 1 February for locations in the southern hemisphere. Day. Whole number.	PCASE	228
STDEV_STOP	Standard deviation of the freezing-season stop day. Day. Tenths.	PCASE	15.5
LATE_STOP	Latest stop day of the freezing season determined over the period of record counted from 1 August for locations in the northern hemisphere or from 1 February for locations in the southern hemisphere. Stop day is when the daily average air temperature rises above the freezing point of water. Day.	PCASE	266
EARLY_STOP	Earliest stop day of the freezing season determined over the period of record counted from 1 August for locations in the northern hemisphere or from 1 February for locations in the southern hemisphere. Stop day is when the daily average air temperature rises above the freezing point of water. Day.	PCASE	198
MLFS	Average of the number of days in the freezing season over the period of record. Day. Tenths.	PCASE	124.8
STDEV_LFS	Standard deviation of the length of the freezing season in days Day. Tenths.	PCASE	20.3
YRLY_AVG_HDD	Average annual maximum cumulative heating degree-days.[1] Celsius degree-days. Tenths	PCASE	3326.8
YRLY_STDEV_HDD	Standard deviation of annual maximum cumulative heating degree-days.[1] Celsius degrees-days. Tenths.	PCASE	219.3
HDD_PCT95	95th percentile of the annual maximum heating degree-days.[1] Celsius degrees-days. Tenths.	PCASE	3765.7
HDD_PCT90	90th percentile of the annual maximum heating degree-days.[1] Celsius degrees-days. Tenths.	PCASE	3709.4

WorldIndex Parameter Identifier Used as Column Header	Definition	Type	Minot, ND, Value
HDD_PCT50	50th percentile of the annual maximum heating degree-days.[1] Celsius degrees-days. Tenths.	PCASE	3324.6
HDD_PCT10	10th percentile of the annual maximum heating degree-days.[1] Celsius degrees-days. Tenths.	PCASE	3050.0
HDD_PCT05	5th percentile of the annual maximum heating degree-days.[1] Celsius degrees-day .Tenths.	PCASE	2989.8
HDD_PCT03	3rd percentile of the annual maximum heating degree-days.[1] Celsius degrees-days. Tenths.	PCASE	2989.8
YRS_FT	Number of years of data used to calculate the number of freeze–thaw cycles. Years.	PCASE	37
FT_AVG	Average number of annual freeze–thaw cycles over the period of record. A freeze–thaw cycle, as reported in WorldIndex, occurs when the daily average air temperature goes above the freezing point of water after being below freezing the prior day. Unitless.	PCASE	13.4
FT_STDEV	Standard deviation of the number of freeze–thaw cycles over the period of record. Unitless. Tenths.	PCASE	3.8
JAN_AVG_TEMP FEB_AVG_TEMP MAR_AVG_TEMP APR_AVG_TEMP MAY_AVG_TEMP JUN_AVG_TEMP JUL_AVG_TEMP AUG_AVG_TEMP SEP_AVG_TEMP OCT_AVG_TEMP NOV_AVG_TEMP DEC_AVG_TEMP	Average temperature of each month. Degrees Celsius. Tenths.	PCASE	−10.1 −7.7 −1.3 6.9 13.5 18.5 21.7 21.1 15.5 7.8 −1.1 −8.6

WorldIndex Parameter Identifier Used as Column Header	Definition	Type	Minot, ND, Value
JAN_MAX_TEMP FEB_MAX_TEMP MAR_MAX_TEMP APR_MAX_TEMP MAY_MAX_TEMP JUN_MAX_TEMP JUL_MAX_TEMP AUG_MAX_TEMP SEP_MAX_TEMP OCT_MAX_TEMP NOV_MAX_TEMP DEC_MAX_TEMP	Average maximum temperatures of each month. Degrees Celsius. Tenths.	PCASE	−5.3 −2.7 4.0 13.6 20.4 25.1 28.6 28.5 22.6 13.9 3.8 −3.9
JAN_MIN_TEMP FEB_MIN_TEMP MAR_MIN_TEMP APR_MIN_TEMP MAY_MIN_TEMP JUN_MIN_TEMP JUL_MIN_TEMP AUG_MIN_TEMP SEP_MIN_TEMP OCT_MIN_TEMP NOV_MIN_TEMP DEC_MIN_TEMP	Average minimum daily temperature of each month. Degrees Celsius. Tenths.	PCASE	−15.0 −12.6 −6.6 0.1 6.5 12.0 14.8 13.6 8.4 1.7 −6.1 −13.3
REQ START DATE	Beginning of the analysis period. DDMMMYYYY. (Note this format implies 2 digits for the date (DD), three alpha characters for the month (MMM), and four digits for the year (YYYY). See example in column 4 of this line.)	IR	01AUG1980
REQ END DATE	End of the analysis period. DDMMMYYYY.	IR	31JUL2017
ACTUAL START DATE	First date for which data was found. ACTUAL START DATE is the same as or later than the REQ START DATE. DDMMMYYYY.	IR	01AUG1980
ACTUAL END DATE	Last date for which actual data was found. ACTUAL END DATE is the same as or earlier than the REQ END DATE. DDMMMYYYY.	IR	31JUL2017
NOMINAL DATA	Number of days between the ACTUAL START DATE and ACTUAL END DATE, inclusive. Days.	IR	13514

WorldIndex Parameter Identifier Used as Column Header	Definition	Type	Minot, ND, Value
ACTUAL DATA	Number of days with data between the ACTUAL START DATE and ACTUAL END DATE, inclusive. Days.	IR	13501
CI	Difference between the average annual temperature of the warmest year (1 January through 31 December) and the coldest year. Parameter used to identify stations with bad data. Degrees Celsius.	IR	5.15
DTRATIO_MAX	Maximum ratio of the difference between the monthly average temperature and monthly minimum temperature to the difference between the monthly maximum temperature and the monthly minimum temperature. Selected from the ratio calculated for each of the 12 months of the year. Parameter used to identify stations with bad data. Unitless.	IR	0.5
DTRATIO_MIN	Minimum ratio of the difference between the monthly average temperature and monthly minimum temperature to the difference between the monthly maximum temperature and the monthly minimum temperature. Selected from the ratio calculated for each of the 12 months of the year. Parameter used to identify stations with bad data. Unitless.	IR	0.5

[1] In this context, a heating degree-day is equivalent to a thawing degree-day.

3 Data Sources

Historical climate data used to update the WorldIndex database came from two sources. The first source was the Global Surface Summary of Day (GSOD) data version 7 from the National Centers for Environmental Information (NCEI) (NOAA [National Oceanic and Atmospheric Administration] 2018a). This database consists of daily summaries of the data recorded at stations worldwide. GSOD data is based on synoptic or hourly station observations that are available in the Integrated Surface Database (ISD) (NOAA 2018b). The available historical data begins in 1929, but the most complete data set is considered to cover 1973 to the present.

Each station included in the GSOD data set is identified by up to three pieces of data: a U.S. Air Force / World Meteorological Organization (USAF/WMO) number, a Weather Bureau Army Navy (WBAN) number, and a station name. Not all the stations have all three identifiers. The stations are grouped by country and in the U.S. by state. The station metadata includes latitude, longitude, and elevation. In general, stations outside the U.S. have a USAF/WMO number but not a WBAN number. In these cases, the WBAN number is assigned 99999 to indicate a missing value. Stations located in the U.S. generally have a WBAN number, but many do not have a USAF/WMO number. In these cases, the USAF/WMO number is assigned 99999 to indicate a missing value.

The GSOD data set is available through FTP download from the NCEI. All the GSOD data for a given year are stored in a single directory. Files are identified by a hyphenated combination of the USAF/WMO number and the WBAN number followed by the year. Given this data-storage structure, it was not possible to download in one step all the data for a station covering the complete period of interest. As a result, the data for each station was downloaded year by year to cover 1 August 1980 through 31 July 2017. The CRREL database, as described in the next section, was then used to assemble the available data for each station into a correctly ordered time series.

NOAA provides a brief review of problems and issues with the ISD data set (NOAA 2018c, 2018d). These are often the result of problems in the original reported or transmitted data or with the metadata. NOAA applies corrections as time and resources permit. As described in the next section, data quality checks during the WorldIndex update identified and removed stations with problems.

The second data source for the WorldIndex update was the Global Historical Climatology Network (GHCN) (Menne et al. 2012a). This database includes data from 30 sources. Only data provided by the National Weather Service Cooperative Observer Program (COOP) was used in this update.

Each COOP station included in the GHCN data set is identified by a GHCN ID beginning with "USC" and followed by its COOP identification number. The first two numbers of the COOP identification number identify the state where the station is located. Each station also has a name. The station metadata include latitude, longitude, and elevation.

The GHCN data set is available through FTP download from NCEI. All the GHCN data for all stations for all years are stored in a single directory, which is available for download as a single archive file.

Menne et al. (2012b) describes the quality assurance procedures applied to the GHCN data set. The multitiered quality assurance consisted primarily of routine, fully automated procedures with some additional overall data record integrity checks. The automated procedure consists of format checking and a comprehensive sequence that identifies daily values that violate any of 19 quality tests. This system flags approximately 0.3% of over 2 billion data values. Menne et al (2012b) estimate that 98%–99% of the values flagged are true data errors and only 1% to 2% are false positives (i.e., flagged as bad in error).

4 WorldIndex Update Data Processing

The WorldIndex update data processing required three major components. The first component was acquiring the data from NCEI and loading the data in the CRREL database. The CRREL database, a preliminary step in the update of the WorldIndex database, was the U.S. Army Corps of Engineers (USACE) Hydrologic Engineering Center Data Storage System (HEC-DSS) (USACE 2009, 1991, 1987). The second component was preliminary processing of the NCEI data and merging files that belong to the same station. The third was the final processing of the NCEI data to produce the updated WorldIndex database.

Several computer programs and a number of scripts were developed to support the WorldIndex update. In general, the scripts were used to transfer data and to automatically run the processing program. The major processing program developed in this project was XICE, which reads the daily maximum and minimum air temperatures from the HEC-DSS database and calculates the parameters for the WorldIndex database for each station.

4.1 Data acquisition and entering data in the CRREL database

As described above, the historical climate data used to update the WorldIndex database came from two sources, both of which were available through FTP download from NCEI. All of the data was entered in the CRREL HEC-DSS database as a preliminary step. HEC-DSS is available as a free download from the Hydrologic Engineering Center. It is an optimal database for storing and retrieving large sets of time-series data. As all of the air-temperature data downloaded from NCEI was time-series data, HEC-DSS was well-adapted to this application.

HEC-DSS uses a block of sequential data as the basic unit of storage. Data is stored in blocks, or records, within a file; and each record is identified by a unique name called a “pathname.” A pathname is separated into six parts, which may be up to 32 characters each. The pathname parts are labeled “A” through “F.” Parts “C,” “D,” and “E” are reserved for data parameter description, starting date of time block, and the time interval at which the data is collected, respectively. This leaves the “A,” “B,” and “F” parts available to provide a unique identifier for each station.

Because the GSOD data and the GHCN data are stored in different formats on NCEI, they required slightly different processes for data acquisition and loading data into the CRREL database.

4.1.1 GSOD data

As described in section 3, all the GSOD data for a given year is stored in a single directory at NCEI in a compressed file. This compressed year file was downloaded to CRREL using FTP and uncompressed. The uncompressed file contained separate files for each station for that year.

The station files were processed in the order that they were listed in the GSOD station list (ftp://ftp.ncdc.noaa.gov/pub/data/noaa/isd-history.txt). (There are 29,820 stations in this list; however, not all are unique.) First, the GSOD station list was processed sequentially to find the USAF/WMO and WBAN number of each station. Next, the USAF/WMO and WBAN numbers were used to create the file name for that station for the given year. The station files were then searched to locate the data file with a matching name. If the data file was found, it was opened and the data written to the CRREL database. This process was repeated until all the stations in the GSOD station list had been processed for that year. Then the compressed file for the next year was downloaded and the process repeated.

The country name, the station name, and the USAF/WMO number were entered into the "A," "B," and "F" pathname parts of the CRREL database, respectively. The state name was entered into the "A" pathname part for stations in the U.S. Outside of the U.S., the combination of the USAF/WMO number and station name was unique for every station. In the U.S., there were roughly 140 stations that shared a USAF/WMO number and a station name with another station. In these cases, almost all the stations had unique WBAN numbers. However, the WBAN number was not used as part of the pathname. As a result, stations that shared a USAF/WMO number and a station name were merged in the CRREL database. In all cases, these were identical stations that happened to have a change in the WBAN number at some point in time.

4.1.2 GHCN (COOP) data

As described in section 3, the GHCN data for all stations for all years is stored in a single file that is available for download as a .tar file. This compressed file was downloaded to CRREL using FTP and uncompressed. The uncompressed file contains separate files for each station. Data for each COOP station is identified by "USC" and a unique COOP number. The GHCN station list (https://www1.ncdc.noaa.gov/pub/data/ghcn/daily/ghcnd-stations.txt) includes number, name, and metadata for each COOP station. There were 22,474 COOP stations in the GHCN station list on 18 October 2018.

Next, the GHCN station inventory (https://www1.ncdc.noaa.gov/pub/data/ghcn/daily/ghcnd-inventory.txt) was reviewed, and the 9500 COOP stations with daily air-temperature maximums for any years from 1975 to the present were selected. This information was used to create the appropriate file name for each station. The station files were then searched, and files with GHCN numbers matching the COOP inventory list were opened and the data written to the CRREL database. This process was repeated until all the chosen COOP stations had been processed.

The state name, the station name, and the COOP number were entered into the "A," "B," and "F" pathname parts of the CRREL database, respectively. There is a unique COOP number for each station.

4.2 Preliminary processing and merging of duplicate stations

The GSOD data required some preliminary processing to merge files that were from the same station before that data could be analyzed for the WorldIndex update. The multiple files for a single station occur for a variety of reasons:

- USAF/WMO numbers may be 99999 for some of the station files and specified for others.
- WBAN numbers may be 99999 for some of the station files and specified for others.
- USAF/WMO or WBAN numbers may change.
- The station name may be different because of a change in abbreviation, change in word order, or change in airport name, for example.
- The longitude, latitude, or elevation may change because of increasing accuracy and precision over time; a station location may change; values may not have been known; or a units became mixed up.
    - Older longitudes and latitudes were reported to 0.1° while newer values were to 0.001°.
    - Some were converted to decimal degrees from degrees and minutes with the precision of the minutes value varying over time (e.g., nearest 10, 5, or 1 minute).

- Station location may change as runways are added at an airport.
- Missing station elevations may be reported as zero.
- Elevation conversions from feet to meters may have been erroneous.

The periods of record for these multiple versions of GSOD data from a single station may be the same, may overlap, or may be completely different. Table 3 shows three examples of groups of stations that are the same or are close to each other.

Table 3. Examples of stations in the GSOD database with similar latitudes and longitudes.

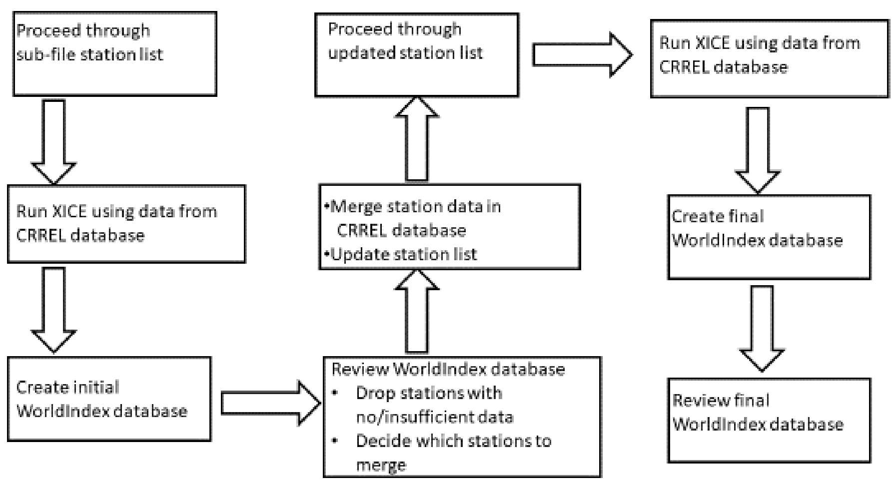
USAF/WMO	WBAN	Station Name	ST	LAT	LONG	ELEV	Start	End
747540	13934	ALEXANDRIA INT	LA	31.317	−92.55	27	19430901	19991231
747540	99999	ALEXANDRIA INT	LA	31.317	−92.55	27	20000101	20051231
747540	93915	ALEXANDRIA INTL AIRPORT	LA	31.335	−92.559	25.6	20060101	20160502
999999	93915	ALEXANDRIA INTL AIRPORT	LA	31.335	−92.559	24.4	20011001	20051231
720272	94282	SKAGIT REGIONAL AIRPORT	WA	48.467	−122.417	42.7	20060101	20160502
720272	99999	SKAGIT RGNL	WA	48.467	−122.417	44	20040525	20051231
998007	99999	PADILLA BAY RESERVE	WA	48.467	−122.467	3	20071016	20160501
742006	99999	BURLINGTON MT VERN	WA	48.47	−122.42	43	19850618	20000104
604810	99999	ORAN-TAFARAOUI	AG	35.483	−0.517	111	19811101	20110706
605054	99999	ORAN TAFAROUI	AG	35.483	−0.533	0	19421114	19851124
604592	99999	TAFARAOUI	AG	35.533	−0.533	115	20040525	20160501

The GSOD station list was divided into 31 separate subfiles for processing convenience. Each subfile list was grouped alphabetically by country or U.S. state. The preliminary processing of the GSOD data and the merging of duplicate stations followed the diagram shown in Figure 1.

Each subfile was run separately. First XICE was run to generate the initial WorldIndex database for the stations in the subfile. Within each subfile list, stations were designated to be merged if their latitudes and longitudes differed by less than 0.051°. The records of up to six stations could be combined if they met these criteria. This eliminated separate versions of data at the same location and obtained longer periods of record. At the same time, stations names containing BOGUS, BUOY, PLATFORM, JETTY, _DOCK, or LIGHTSHIP were used to identify bogus or offshore stations for removal. Stations with no data were also designated for removal. The final result of the merging procedure was a new subfile station listing that indicated which stations were to be combined in the CRREL database and which would be removed. The identification of the combined stations used the USAF/WMO number and station name of the station with the longest period of record. If the periods of record were equal for two or more stations, then the USAF/WMO number and station name of the station with the smallest USAF/WMO number was used.

The merging procedure was used to remove stations with no data and to combine duplicate stations in the CRREL database. XICE was then run based on the revised subfile lists. A final WorldIndex database was created for each subfile. From the 26,572 GSOD stations downloaded, 16,008 unique and possibly useful locations with at least some data were included in the WorldIndex database.

Figure 1. Preliminary processing and merging duplicate stations.



4.3 Updated WorldIndex database

Some data corrections were made in the WorldIndex in postprocessing:

- Removed a station in Afghanistan with a latitude and longitude located in California.
- Removed three rogue stations in the U.S. with a mismatch in state, station name, and location.
- Removed 104 stations with Condition Index (CI) (see Table 2) values greater than 11.6. (Large values of the internal review parameter CI, discussed below, indicate systematic errors in the air-temperature record). The value of 11.6 was chosen by reviewing the temperature time series of six stations in the U.S. with high CI values. Most (85) of the stations rejected are in Canada. All these station appear to have many

incorrect temperatures apparently because of a mix-up between Fahrenheit and Celsius when the data was archived in GSOD.

The final count of global stations in the database with at least 5 years of PCASE-quality data was 16,668 with 8868 of them in the U.S.; and of those, GHCN provided 6576 COOP stations. Figures 2 and 3 show the locations of stations in the updated WorldIndex database. Figure 4 compares the periods of record of these stations to those in the previous WorldIndex database. Figure 5 compares the distribution of station location by latitude and longitude to the previous WorldIndex.

Figure 2. WorldIndex stations: western hemisphere.

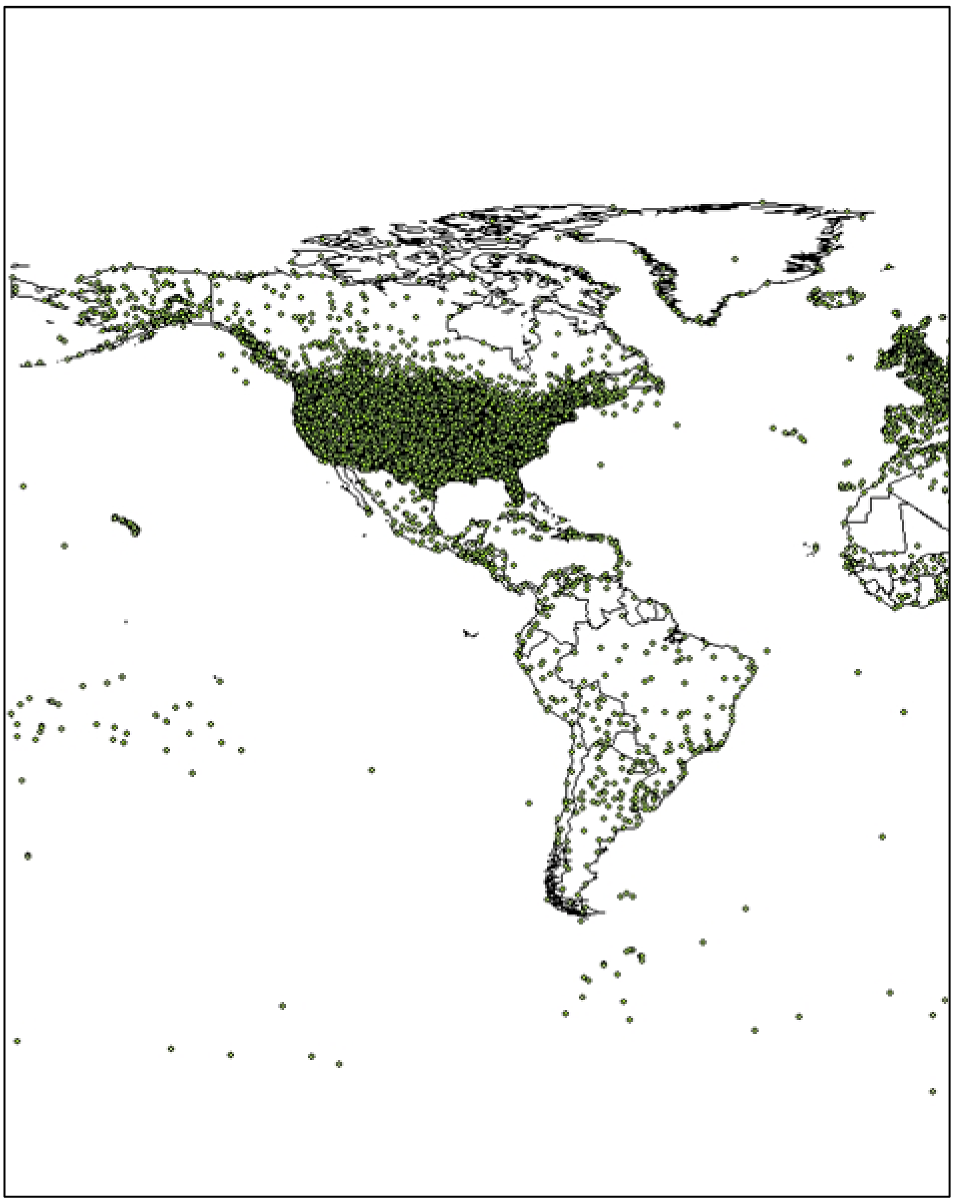


Figure 3. WorldIndex stations: eastern hemisphere.

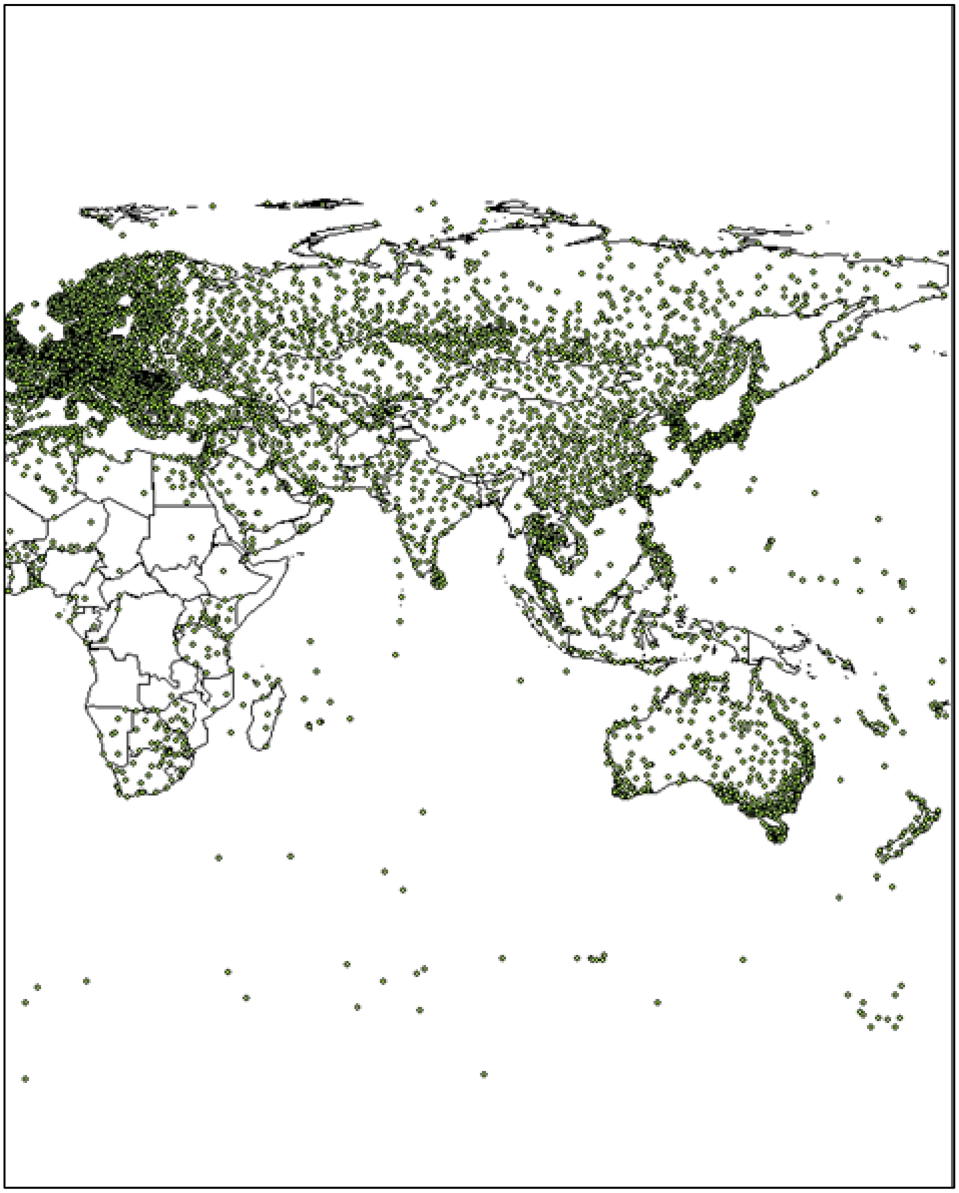


Figure 4. Periods of record for previous WorldIndex stations (*top*) and updated WorldIndex stations (*bottom*).

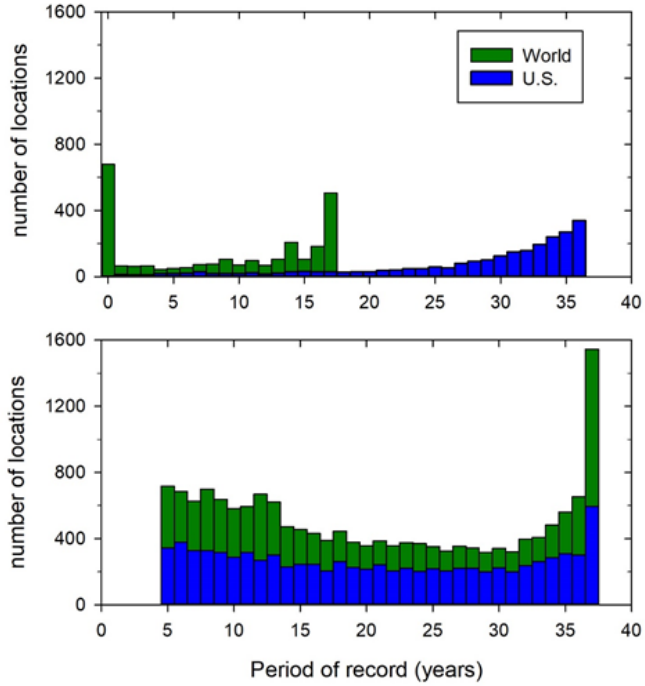
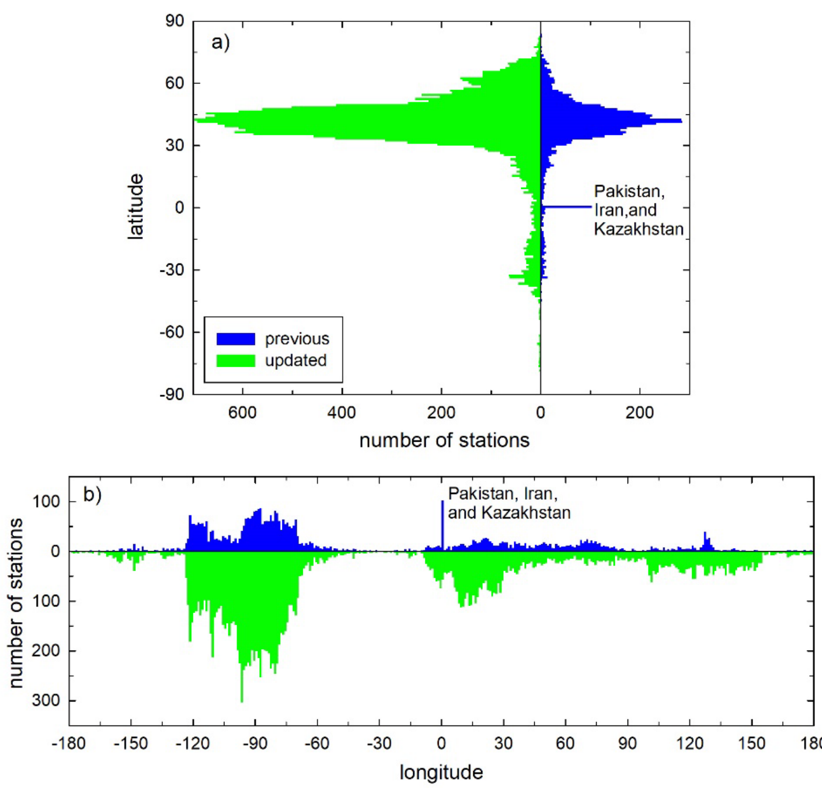


Figure 5. Distribution of stations by latitude (*top*) and longitude (*bottom*) for previous and updated WorldIndex. The spike at zero latitude and longitude in the previous WorldIndex indicates stations without latitude and longitude values.



5 Continuity of WorldIndex Database Stations

This section describes the continuity and consistency of WorldIndex database stations between the previous and updated versions. *Continuity* describes the number of stations in the previous database that appear in the updated version. Ideally, weather stations would never be discontinued so that any update of WorldIndex would include the stations from the previous version of the database and any new stations established in the intervening years. However, stations are sometimes abandoned, or countries may not report station data that was reported in the past. *Consistency* describes the similarity of the data values in the previous and updated versions. The consistency of the update is reviewed using two representative WorldIndex parameters: yearly average temperature (YRLY_TAVG) and the yearly average freezing degree-days (YRLY_AVG_FDD). (See Table 2 for the definitions of YRLY_TAVG and YRLY_AVG_FDD).

Three categories describe the continuity of stations included in the WorldIndex data. "Continuing" describes stations that have the same name and code in the previous and updated version of the WorldIndex database and are at the same location. "Replaced" describes stations that do not have matching names and codes in the previous and updated version, but a station exists in the updated version that is located less than 10 km away. "Deleted" describes stations that do not have the same name and code in the updated version and no station located within 10 km.

There were 2497 stations located in the U. S. in the previous WorldIndex database. 1898 of these stations are continuing in the updated database, 487 are replaced, and 112 are deleted. There were 2269 stations in the previous WorldIndex database located outside of the U. S. 1706 of these stations are continuing in the updated database, 98 stations are replaced, and 465 stations are deleted.

A comparison of the yearly average temperature (YRLY_TAVG) and the yearly average freezing degree-days (YRLY_AVG_FDD) was used to review the consistency of the previous and updated WorldIndex databases. Only continuing stations with data were included in the comparison. Figure 6 shows the results for the U.S. In these charts, the updated WorldIndex results are plotted on the y-axis and the previous WorldIndex results are plotted along the x-axis. The values are close, with the average updated yearly average temperature 0.3°C higher and the average updated yearly average freezing degree-days 41 °C-days lower on average than the previous values. Figure 7 compares stations outside of the U.S. Similar to the U.S., the values show little change, with the average updated yearly average temperature 0.5°C higher and the average updated yearly average freezing degree-days 20 °C-days lower than the previous average values.

Figure 6. Comparison of yearly average temperature and yearly average freezing degree-days between the previous and updated WorldIndex databases for the U.S.

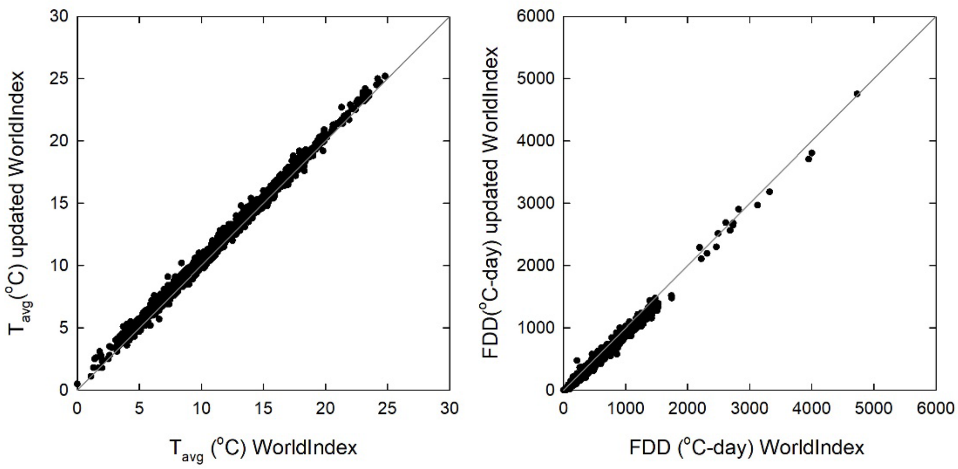
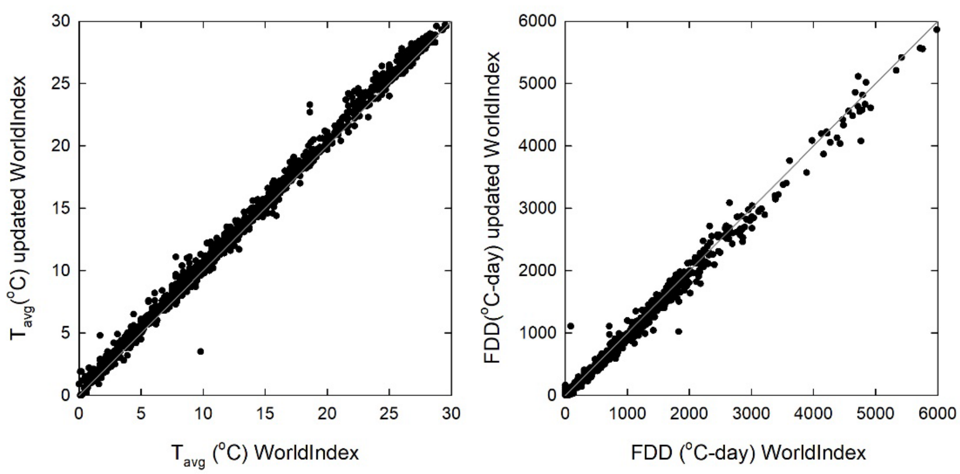


Figure 7. Comparison of yearly average temperature and yearly average freezing degree-days between the previous and updated WorldIndex databases outside of the U.S.



6 Determination of the WorldIndex Parameters

This section describes the steps used by XICE to determine the temperature parameters required by the WorldIndex database. Variables are shown in italics, arrays in bold, and the WorldIndex parameters in all capitals that match their listing in Table 2. The first day of the year is 1 August in the northern hemisphere and 1 February in the southern hemisphere. (As discussed in Appendix A, these starting points were chosen to occur before the onset of subfreezing winter temperatures to allow proper calculation of the freezing degree-days parameter.)

6.1 Data review and initial data analysis

6.1.1 Data review

Two one-dimensional data arrays are used to estimate all values provided for PCASE. These data arrays are $\mathbf{T_{max}}$ and $\mathbf{T_{min}}$, where $\mathbf{T_{max}}$ contains the maximum daily air temperatures listed in consecutive order (with missing value markers for days with missing values) and $\mathbf{T_{min}}$ contains the minimum daily air temperatures listed in consecutive order (with missing value markers for days with missing values). The lengths of the $\mathbf{T_{max}}$ and $\mathbf{T_{min}}$ arrays are equal. Their nominal length is determined by the initial start and end dates requested from the HEC-DSS database; however, their actual length may be shorter. The two data reviews, the Missing Data Test and the Data Sufficiency Test, are based on the $\mathbf{T_{max}}$ and $\mathbf{T_{min}}$ arrays:

- **Missing Data Test**. If the $\mathbf{T_{max}}$ and $\mathbf{T_{min}}$ arrays are empty, that is they contain zero days with actual data, the station data set is considered missing; and all the WorldIndex parameters are set to 88888 (Table 4).
- **Data Sufficiency Test**. The station arrays are inspected to make sure there are at least 5 years with at least 350 days of data. If the station fails this test, all the WorldIndex parameters are set to 99999 (Table 4).

Table 4. Initial data review status indicators.

Condition	Output Parameter values
No data	**SD:** As is **PCASE:** 8888 **IR:** 8888 except REQ_START_DATE and REQ_END_DATE as is; ACTUAL_START_DATE and ACTUAL_END_DATE: "xxxx"
Years < 5 with days of data ≥ 350	**SD:** As is **PCASE:** 9999 **IR:** 9999 except REQ_START_DATE and REQ_END_DATE as is; ACTUAL_START_DATE and ACTUAL_END_DATE as determined by available data.

6.1.2 Climatology

The hemisphere of the globe (northern or southern) is determined by the latitude of the station. Stations in the northern hemisphere have a positive latitude, and stations in the southern hemisphere have a negative latitude.

The preliminary XICE data analysis cycles through all the available data from the date of the first data point, JD_1, through the date of the final data point, JD_{max}. JD is the Julian date as defined by the HEC-DSS database as days since 31 December 1899. The number of potential days with data, N, is found as $N = JD_{max} - JD_1 + 1$.

The station data are converted to the required engineering units. WorldIndex requires results in SI units (°C), the GSOD data are in customary units (°F), and GHCN data are in SI units (°C).

The daily average air temperature, $\mathbf{T_{avg}}$, is determined for each day as

$$\mathbf{T_{avg}}(j) = \frac{\mathbf{T_{max}}(j) + \mathbf{T_{min}}(j)}{2} \text{ for } j = 1, N.$$

If either $\mathbf{T_{max}}$ or $\mathbf{T_{min}}$ is missing on any day j, then $\mathbf{T_{avg}}$ is set missing for that day.

Next, the climatology for the daily average temperature is determined from the available nonmissing data. The climatology is developed for $\mathbf{T_{avg}}$ for each Julian day of the year, J_{DOY}, where $1 \le J_{DOY} \le 365$. The climatology analysis cycles through all $\mathbf{T_{avg}}$ (j) from the date of the first data point, JD_1, to the date of the final data point, JD_{max}. On each day, the Julian date as defined by the HEC-DSS database, *JD*, is converted to the Julian Day of Year, J_{DOY}. Then the climatology estimate of $\mathbf{T_{avg}}$ is found for each J_{DOY} as

$$\mathbf{T_{avgDOY}}\left(J_{DOY}\right)=\frac{1}{N_{DOY}}\sum_{j=1}^{N_{DOY}}\mathbf{T_{avg}}\left(JD=J_{DOY}\right),$$

where $\mathbf{T_{avgDOY}}$ (J_{DOY}) is the average daily air temperature averaged on that Julian day of year over all the years in the period of record and N_{DOY} is the number of years with nonmissing data on J_{DOY}.

Then the data analysis cycles through all $\mathbf{T_{avg}}$ (j) from the date of the first data point, JD_1, to the date of the final data point, JD_{max}. $\mathbf{T_{avgDOY}}$ ($j = J_{DOY}$) is substituted for $\mathbf{T_{avg}}$ (j) for any day with missing data. At this point, $\mathbf{T_{avg}}$ (j) should have no missing data. (However, if there was no data recorded on a particular day of year for all the years of record, then it is possible for $\mathbf{T_{avg}}$ (j) to have missing days.

6.1.3 Cumulative degree-days

The data analysis cycles through all $\mathbf{T_{avg}}$ (j) from the date of the first data point, JD_1, to the date of the final data point, JD_{max} to calculate the cumulative degree-days for each day. (See Appendix A for a discussion of the cumulative degree-days.) The degree-days accumulation begins on the first day of the year, J_{DOY} =1, and continues through J_{DOY} = 365 for every year.

The cumulative degree-days for each day, **CDD** (j), is determined using the following procedure: For all j from the date of the first data point, JD_1, to the date of the final data point, JD_{max},

$$\text{if } J_{DOY}=1, \mathbf{CDD}\left(j\right)=\left(\mathbf{T_{avg}}(j)-T_{base}\right),$$

$$\text{if } J_{DOY}>1,\ \mathbf{CDD}\left(j\right)=\left(\mathbf{T_{avg}}(j)-T_{base}\right)+\mathbf{CDD}\left(j-1\right),$$

where T_{base} = 0°C in SI units and 32°F for customary units. It can be seen that **CDD** (j) is reset on the first day of the year. The **CDD** (j) on J_{DOY} =365 is retained as an End-of-Year value, **CDD_EOY** (*year*).

6.1.4 Monthly average temperatures

The data analysis cycles through all $\mathbf{T_{avg}}$ (*j*) from the date of the first data point, JD_1, to the date of the final data point, JD_{max}, to calculate the monthly average temperatures and the yearly temperatures. The daily average temperature, daily maximum temperature, and daily minimum temperature are summed for each month of the year. The daily average temperature is also summed for each year. Later, these sums will be used to determine the monthly average temperature, **Month_avg** (*imon*); monthly average maximum temperature, **Month_max** (*imon*); and monthly average minimum temperature, **Month_min** (*imon*), where *imon* is the month of the data (1–12), as well as the yearly average temperature, **Yearly_avg** (*year*).

For all *j* from the date of the first data point, JD_1, to the date of the final data point, JD_{max}, the year, month (1–12), and day (1–31) of the month are determined based on *j*. All variables that end as *_sum* or *_count* are initialized with a value of zero at the start.

$$\mathbf{Month\_sum}(imon) = \mathbf{Month\_sum}(imon) + \mathbf{T}_{\mathbf{avg}}(j)$$
$$\mathbf{Month\_count}(imon) = \mathbf{Month\_count}(imon) + 1$$
$$\mathbf{Month\_max\_sum}(imon) = \mathbf{Month\_max\_sum}(imon) + \mathbf{T}_{\mathbf{max}}(j)$$
$$\mathbf{Month\_min\_sum}(imon) = \mathbf{Month\_min\_sum}(imon) + \mathbf{T}_{\mathbf{min}}(j)$$
$$\mathbf{Year\_avg\_sum}(year) = \mathbf{Year\_avg\_sum}(year) + \mathbf{T}_{\mathbf{avg}}(j)$$
$$\mathbf{Year\_count}(year) = \mathbf{Year\_count}(year) + 1$$

As this calculation uses only nonmissing data, **Month_count** (*imon*) is the number of nonmissing data in each month, and **Year_count** (*year*) is the number of nonmissing data in each year.

The end of a freeze–thaw cycle is defined by $\mathbf{T_{avg}}$ (*j*) > T_{base} and $\mathbf{T_{avg}}$ (*j-1*) ≤ T_{base}. The number of times this occurs in a year is counted and entered into **ANN_FT_cycle** (*year*) on 30 September. If no freeze–thaw cycle occurs, then the number of times is 0.

6.2 Analysis

6.2.1 YRLY_TAVG, YRLY_STDEV_TAVG, YRS_TAVG

The yearly average temperature is first found for each year as

$$\textbf{Yearly\_avg}(year) = \textbf{Year\_avg\_sum}(year) / \textbf{Year\_count}(year).$$

Then the WorldIndex parameters are found as follows:

$$\text{YRLY\_TAVG} = \frac{1}{N_{year}} \sum_{j=1}^{N_{year}} \textbf{Yearly\_avg}(j),$$

$$\text{YRLY\_STDEV\_TAVG} = \sqrt{\frac{1}{(N_{year}-1)} \sum_{j=1}^{N_{year}} \left(\textbf{Yearly\_avg}(j) - \text{YRLY\_TAVG}\right)^2},$$

$$\text{YRS\_TAVG} = N_{year},$$

where N_{year} is the number of years.

6.2.2 YRLY_AVG_FDD, YRLY_STDEV_FDD, YRS_FDD

Refer to Appendix A for a discussion of determining $\mathbf{CDD_{max}}$ (*year*), $\mathbf{CDD_{min}}$ (*year*), $\mathbf{j_{max}}$ (*year*), and $\mathbf{j_{min}}$ (*year*). The freezing degree-days (FDD) for the year, **FDD** (*year*), is calculated as

$$\textbf{FDD}(year) = \textbf{CDD}_{\textbf{max}}(year) - \textbf{CDD}_{\textbf{min}}(year).$$

If no day has $\mathbf{T_{avg}}$ less than or equal to T_{base}, then **FDD** (*year*) =0, $\mathbf{CDD_{max}}$ (*year*) = 0, $\mathbf{CDD_{min}}$ (*year*) = 0, $\mathbf{j_{max}}$ (*year*) = day of year of 01OCT, and $\mathbf{j_{min}}$ (*year*) = day of year of 01OCT.

At this point, the number of years with nonmissing **FDD** (*year*) is counted and assigned to YRS_FDD. If there are no nonmissing years, then YRS_FDD = 9999. If YRS_FDD is not equal to 9999, then the average of **FDD** (*year*), YRLY_AVG_FDD, and the standard deviation of **FDD** (*year*), YRLY_STDEV_FDD, are calculated as

$$\text{YRLY\_AVG\_FDD} = \frac{1}{\text{YRS\_FDD}} \sum_{j=1}^{YRS\_FDD} \textbf{FDD}(j),$$

$$\text{YRLY\_STDEV\_FDD} = \sqrt{\frac{1}{(\text{YRS\_FDD}-1)} \sum_{j=1}^{YRS\_FDD} \left(\textbf{FDD}(j) - \text{YRLY\_AVG\_FDD}\right)^2}.$$

These calculations include only nonmissing years.

6.2.3 FDD PCT95, FDD PCT90, FDD PCT50, FDD PCT10, FDD PCT05, FDD PCT03

The freezing degree-day percentiles are calculated as follows. First, using the procedure ***indexx***, described in Press et al. (1992), **FDD** (*year*) is ranked and placed into a new array, **FDD_ranked** (*k*). The new array has the property that **FDD_ranked** (1) ≤ **FDD_ranked** (2) ≤ …. **FDD_ranked** (YRS_FDD). Then

FDD PCT95 = **FDD_ranked** (***upinteger*** (0.95*YRS_FDD))

FDD PCT90 = **FDD_ranked** (***upinteger*** (0.90*YRS_FDD))

FDD PCT50 = **FDD_ranked** (***upinteger*** (0.50*YRS_FDD))

FDD PCT10 = **FDD_ranked** (***upinteger*** (0.10*YRS_FDD))

FDD PCT05 = **FDD_ranked** (***upinteger*** (0.05*YRS_FDD))

FDD PCT03 = **FDD_ranked** (***upinteger*** (0.03*YRS_FDD))

where the ***upinteger*** function rounds the value in the argument up to the next higher integer value and does not allow the value to be greater than YRS_FDD. If YRS_FDD = 9999 (there are no nonmissing years), then all the freezing degree-day percentiles are set equal to 9999.

6.2.4 HI_TMAX, LO_TMIN

This is the maximum and minimum air-temperature values over the period of record.

HI_TMAX is equal to the maximum value in the array $\mathbf{T_{max}}$ (j) for j = 1 to N.

LO_TMIN is equal to the minimum value in the array $\mathbf{T_{min}}$ (j) for j = 1 to N.

6.2.5 YRS FREEZE, AVERAGE START, STDEV START, LATE START, EARLY START, AVERAGE STOP, STDEV STOP, LATE STOP, EARLY STOP, MLFS, STDEV LFS

The first step is calculating the length of the freezing season each winter in days, **FS_length** (*year*). This is found as

$$\textbf{FS\_length}\ (year) = \mathbf{j_{min}}\ (year) - \mathbf{j_{max}}\ (year)$$

for each year of the record. If $\mathbf{j_{min}}$ (*year*) or $\mathbf{j_{max}}$ (*year*) is missing, then **FS_length** (*year*) is set to missing.

At this point, the number of years with nonmissing **FS_length** (*year*) is counted and assigned to YRS_FREEZE. If there are no nonmissing years, then YRS_FREEZE = 9999.

The number of days between the start of the freezing season and the first day of the freezing year is found as

$$\textbf{Ann\_Start}\ (year) = \mathbf{j_{max}}\ (year) - \textbf{StartofYear}\ (year) + 1,$$

where **StartofYear** (*year*) is the Julian date of the start of the year for the appropriate year for $\mathbf{j_{max}}$ (*year*). If $\mathbf{j_{max}}$ (*year*) is missing, **Ann_Start** (*year*) is set to missing.

The number of days between the end of the freezing season and the start of the year is found as

$$\textbf{Ann\_End}\ (year) = \mathbf{j_{min}}\ (year) - \textbf{StartofYear}\ (year) + 1,$$

where *StartofYear* (*year*) is the Julian date of the start of the year for the appropriate year for $\mathbf{j_{min}}$ (*year*). If $\mathbf{j_{min}}$ (*year*) is missing, **Ann_End** (*year*) is set to missing.

The average of **Ann_Start** (*year*), AVERAGE_START, and the standard deviation of **Ann_Start** (*year*), STDEV_START, are calculated as

$$\text{AVERAGE\_START} = \frac{1}{\text{YRS\_FREEZE}} \sum_{j=1}^{\text{YRS\_FREEZE}} \textbf{Ann\_Start}(j),$$

$$\text{STDEV\_START} = \sqrt{\frac{1}{(\text{YRS\_FREEZE}-1)} \sum_{j=1}^{\text{YRS\_FREEZE}} \left(\textbf{Ann\_Start}(j) - \text{AVERAGE\_START}\right)^2}\ .$$

These calculations include only nonmissing years.

LATE_START is equal to the integer of the maximum value in the array **Ann_Start** (*year*) for *year* = 1 to YRS_FREEZE.

EARLY_START is equal to the integer of the minimum value in the array **Ann_Start** (*year*) for *year* = 1 to YRS_FREEZE.

The average of **Ann_End** (*year*), AVERAGE_END, and the standard deviation of **Ann_End** (*year*), STDEV_END, are calculated as

$$\text{AVERAGE\_END} = \frac{1}{\text{YRS\_FREEZE}} \sum_{j=1}^{\text{YRS\_FREEZE}} \mathbf{Ann\_End}(j),$$

$$\text{STDEV\_END} = \sqrt{\frac{1}{(\text{YRS\_FREEZE}-1)} \sum_{j=1}^{YRS\_FREEZE} \left(\mathbf{Ann\_End}(j) - \text{AVERAGE\_END}\right)^2} \ .$$

These calculations include only nonmissing years.

LATE_END is equal to the integer of the maximum value in the array **Ann_End** (*year*) for *year* = 1 to YRS_FREEZE.

EARLY_END is equal to the integer of the minimum value in the array **Ann_End** (*year*) for *year* = 1 to YRS_FREEZE.

The Mean Length of the Freezing Season, MLFS, and the standard deviation of the Length of the Freezing Season, STDEV_LFS, are calculated as

$$\text{MLFS} = \frac{1}{\text{YRS\_FREEZE}} \sum_{j=1}^{\text{YRS\_FREEZE}} \mathbf{FS\_length}(j),$$

$$\text{STDEV\_LFS} = \sqrt{\frac{1}{(\text{YRS\_FREEZE}-1)} \sum_{j=1}^{\text{YRS\_FREEZE}} \left(\mathbf{FS\_length}(j) - \text{MLFS}\right)^2} \ .$$

These calculations include only nonmissing years.

If YRS_FREEZE = 0 or YRS_FREEZE = 9999, then AVERAGE_START, STDEV_START, LATE_START, EARLY_START, AVERAGE_STOP, STDEV_STOP, LATE_STOP, and EARLY_STOP are all set to 0.

6.2.6 YRLY AVG HDD, YRLY STDEV HDD, HDD PCT95, HDD PCT90, HDD PCT50, HDD PCT10, HDD PCT05, HDD PCT03

HDD (heating degree-day) has a specific meaning in PCASE, which is the cumulative sum of thawing degree-days for the period $\mathbf{j_{min}}$ (*year*) until $\mathbf{j_{max}}$ (*year+1*):

HDD (*year*) = ***CDD_EOY*** (*year*)-$\mathbf{CDD_{min}}$ (*year*) + $\mathbf{CDD_{max}}$ (*year*+1).

At this point, the number of years with nonmissing **HDD** (*year*) is counted and assigned to *YRS_HDD* (Note: *YRS_HDD* is not a WorldIndex parameter). If there are at least five nonmissing years, then YRLY_AVG_HDD and the standard deviation of **HDD** (*year*), YRLY_STDEV_HDD, are calculated as

$$\text{YRLY\_AVG\_HDD} = \frac{1}{YRS\_HDD} \sum_{j=1}^{YRS\_HDD} \mathbf{HDD}(j),$$

$$\text{YRLY\_STDEV\_HDD} = \sqrt{\frac{1}{(YRS\_HDD - 1)} \sum_{j=1}^{YRS\_HDD} \left(\mathbf{HDD}(j) - \text{YRLY\_AVG\_HDD}\right)^2}\,.$$

These calculations include only nonmissing years.

The heating degree-day percentiles are calculated as follows. First, **HDD** (*year*) is ranked and placed into a new array, **HDD_ranked** (*k*), using the procedure, ***indexx***, described in Press et al. (1992). The new array has the property that **HDD_ranked** (1) ≤ **HDD_ranked** (2) ≤ .... **HDD_ranked** (*YRS_HDD*). Then

HDD PCT95 = **HDD_ranked** (***upinteger*** (0.95*YRS_HDD)),

HDD PCT90 = **HDD_ranked** (***upinteger*** (0.90*YRS_HDD)),

HDD PCT50 = **HDD_ranked** (***upinteger*** (0.50*YRS_HDD)),

HDD PCT10 = **HDD_ranked** (***upinteger*** (0.10*YRS_HDD)),

HDD PCT05 = **HDD_ranked** (***upinteger*** (0.05*YRS_HDD)),

HDD PCT03 = **HDD_ranked** (***upinteger*** (0.03*YRS_HDD)), where the ***upinteger*** function rounds the value in the argument up to the next higher integer value and does not allow the value to be greater than *YRS_HDD*. If *YRS_HDD* = 9999 (there are no nonmissing years), then all the heating degree-day percentiles are set equal to 9999.

6.2.7 YRS_FT, FT_AVG, FT_STDEV

The number nonmissing **ANN_FT_cycle** (*year*) is counted and assigned to YRS_FT. If there are at least five nonmissing years, then the average of **ANN_FT_cycle** (*year*), FT_AVG, and the standard deviation of **ANN_FT_cycle** (*year*), FT_STDEV, are calculated as

$$\mathrm{FT\_AVG} = \frac{1}{\mathrm{YRS\_FT}} \sum_{j=1}^{\mathrm{YRS\_FT}} \mathbf{ANN\_FT\_cycle}(j),$$

$$\mathrm{FT\_STDEV} = \sqrt{\frac{1}{(\mathrm{YRS\_FT}-1)} \sum_{j=1}^{\mathrm{YRS\_FT}} \left(\mathbf{ANN\_FT\_cycle}(j) - \mathrm{FT\_AVG}\right)^2}\ .$$

6.2.8 JAN_AVG_TEMP through DEC_AVG_TEMP; JAN_MAX_TEMP through DEC_MAX_TEMP; JAN_MIN_TEMP through DEC_MIN_TEMP

The following equations are cycled through for *imon* = 1, 12. (January through December). For each month that **Month_count** (*imon*) > 0,

$$\mathbf{Month\_avg}(imon) = \mathbf{Month\_sum}(imon) / \mathbf{Month\_count}(imon),$$

$$\mathbf{Month\_max\_avg}(imon) = \mathbf{Month\_max\_sum}(imon) / \mathbf{Month\_count}(imon),$$

$$\mathbf{Month\_min\_avg}(imon) = \mathbf{Month\_min\_sum}(imon) / \mathbf{Month\_count}(imon).$$

For any month that **Month_count** (*imon*) = 0, the monthly average, maximum, and minimum are set equal to 9999.

6.2.9 CI

The Condition Index, CI, is a parameter designed to detect bad data sets that can result from improper units, systematic instrument error, and other profound problems in data sets. First, the average temperature for each calendar year, *year_avg* (*year*), if there are more than 0 days of data for that year, is found as

$$\mathbf{year\_avg}(year) = year\_avg\_sum(year) / year\_count(year).$$

Next, *Yearly_avg_max* is set equal to the maximum value in the array **year_avg** (*year*), and *Yearly_avg_min* is set equal to the minimum value in the array **year_avg** (*year*). Note that only years with more than 250 days of data are included in the search for the maximum and minimum values. CI is then found as the difference between these values:

$$\text{CI} = Yearly\_avg\_max - Yearly\_avg\_min.$$

In practice, the difference between the maximum yearly average temperature and minimum yearly average temperature should not be too large. Large differences, greater than about 12°C, almost always indicate that there is a problem with the data set. The most common problem found was that part of the data set was apparently entered into the National Centers for Environmental Information database in degrees Celsius and part in Fahrenheit, yet the database description reports the entire data set as being in Fahrenheit.

6.2.10 DTRATIO_MAX and DTRATION_MIN

The Condition Indexes, DTRATIO_MAX and DTRATIO_MIN, are parameters designed to detect bad data sets.

$$\text{DTRATIO\_MAX} = \text{MAX}\left[\frac{month\_avg(imon) - month\_min\_avg(imon)}{month\_max\_avg(imon) - month\_min\_avg(imon)} -\right]$$

for *imon* = 1, 12;

$$\text{DTRATIO\_MIN} = \text{MIN}\left[\frac{month\_avg(imon) - month\_min\_avg(imon)}{month\_max\_avg(imon) - month\_min\_avg(imon)} -\right]$$

for *imon* = 1, 12.

In general, the values of DTRATIO_MAX and DTRATIO_MIN should be between 0 and 1. Stations with bad data may have DTRATIO_MAX > 1 and DTRATIO_MIN < 0. However, stations with a sizeable percentages of missing data may also fall slightly outside of these values due to the use of climatology for missing values.

7 Conclusions and Recommendations

This report documents the update to the WorldIndex database conducted to support the PCASE7 software program. It describes the database parameters and the procedures developed to download and process all of the ground-based air-temperature data. The WorldIndex database now contains parameters for over 10,000 global stations. These values were produced using the daily maximum and minimum air temperatures from 1980 to 2017.

Based on this effort, the team recommends the following with regard to future updates of the WorldIndex database.

7.1 Timing of future updates

Conduct future updates to the database every 5 years as specified in the PCASE criteria (USACE 2001).

7.2 Refinement of the WorldIndex database

Refine the WorldIndex database by removing redundant parameters and renaming parameters to reflect modern usage.

7.2.1 Redundant parameters

The average, minimum, and maximum temperature values for the months of January and July are redundant with the addition of the average monthly temperatures. The redundant parameters are JAN_TAVG _AVERAGE, JAN_TMIN_AVERAGE, JAN_TMAX_AVERAGE, JUL _TAVG_AVERAGE, JUL_TMIN_AVERAGE, and JUL_TMAX-AVERAGE.

7.2.2 Renaming parameters to reflect modern usage

For clarity, change the column headers describing the heating degree-days to thawing degree-days as this is a more accurate description.

7.3 Addition of new parameters to the database

Include climate variables of interest for pavement design and evaluation, such as wind speed, precipitation amount, and snow depth. The GSOD data set can include up to 12 daily parameters. These parameters are air temperature (mean, maximum, and minimum), pressure (sea level and station), dew point, visibility, wind speed (maximum gust and maximum sustained), precipitation, snow depth, and a unique indicator for occurrence of each of the following: Fog, Rain or Drizzle, Snow or Ice Pellets, Hail, Thunder, and Tornado/Funnel Cloud. (A complete list of GSOD variable descriptions is available from NCEI (2010). Keep in mind that not all GSOD stations will have all parameters. In fact, many stations are limited to the air-temperature parameters and precipitation parameter only. The GHCN station files have five core elements: maximum temperature, minimum temperature, precipitation, snowfall, and snow depth. Each GHCN station includes at least one of these elements (Menne et al 2012b). There are many additional GHCN elements describing cloudiness, wind, soil temperature, ice thickness, sunshine, snow water equivalent, and weather type that may be included in a station file.

7.4 Alternatives to surface observations as a source of global climate data

For the entire history of PCASE, global climate data has been based on surface observations made at ground-based stations. This report discusses disadvantages of this approach, including lack of coverage in certain areas of the world, data gaps for specific time periods at many stations, and problems with data accuracy. One alternative to surface observations would be to base the global climate data on *reanalysis data*. Reanalysis is a systematic approach to produce data sets for climate monitoring based on data assimilation and models. Data assimilation describes the process by which the models ingest surface observations, radiosonde, satellite, buoy, aircraft, ship reports, and other data sources to determine the climate state at each model time step. Key strengths of reanalysis include the global data sets, consistent spatial and temporal resolution over thirty years or more, and hundreds of variables available. In addition, reanalysis data sets are relatively straightforward to handle from a processing standpoint (although file sizes can be very large) (Dee et al 2016). Limitations include relatively large pixels sizes and unknown and changing biases due to the changing mix of observations.

A number of reanalysis data sets are available. Most have global coverage although some are limited to specific areas, such as the Arctic. (See Dee et al. 2016 for an overview of 15 different reanalysis data sets.) One global data set that is a potential candidate for use by PCASE is the Modern Era Retrospective analysis for Research and Applications, Version 2 (MERRA-

2), available from the National Aeronautics and Space Administration (NASA 2019; Rienecker et al. 2011; Gelaro et al. 2017) The MERRA-2 data set begins in 1980, offers worldwide coverage, and has a spatial resolution of ½° latitude by ⅝° longitude (Global Modeling and Assimilation Office 2016) that includes daily average air temperatures, a critical parameter used in pavement design and evaluation.

Recent work for the Federal Highway Administration's Long-Term Pavements Program compared predicted pavement performance from the Mechanistic-Empirical Pavement Design Guide (MEPDG) using continuous hourly climate data from MERRA and ground-based observations and concluded that MERRA data is as good as or better than data collected from ground-based observations (Schwartz et al. 2015). One limitation of MERRA-2 data compared to measured values is the application on variable terrain with higher elevations (K. Jones, CRREL, pers. comm., 2019). Schwartz et al. (2015) noted a similar observation when comparing pavement distress predictions from the MEPDG using MERRA data and automatic weather systems where there was better agreement for sites located on either flat terrain or in northern latitudes. Therefore, future updates to the WorldIndex database should consider reanalysis data if worldwide spatial gridded coverage is advantageous.

7.5 Potential for automation of the WorldIndex database update

The update described in this report was largely automated. Manual requirements mostly focused on problems with GSOD data, including duplicate stations, rogue stations, and other issues as described in this report. If the next update was limited to the GHCN database, then automation of the entire process could be possible. Automation would probably be even more straightforward if a reanalysis data set or other global, gridded data set was adopted as the source of global data for PCASE.

PCASE is a crucial tool for pavement design, and accuracy is paramount. Therefore, the update to the WorldIndex database aims to increase the data available, which will ensure more reliable results. Additionally, implementing the above recommendations will further refine the database and continue to enhance its capabilities. This is essential to support the pavement design work that is fundamental across the U.S. military.

References

Adolf, M. 2010. *PCASE 2.09 User Manual*. U.S. Army Corps of Engineers, Transportation Systems Center and Engineering Research and Development Center. https://transportation.erdc.dren.mil/pcase/downloads/PCASE209%20User%20Manual.pdf.

Cortez, E. R., M. A. Kestler, and R. L. Berg. 2000. "Computer-Assisted Calculations of the Depth of Frost Penetration in Pavement-Soil Structures." In *Proceedings of the Transportation Systems 2000 Workshop*, 28 February–3 March, San Antonio, TX.

Dee, D., J. Fasullo, D. Shea, J. Walsh, and National Center for Atmospheric Research Staff, ed. 2016. "The Climate Data Guide: Atmospheric Reanalysis: Overview & Comparison Tables." Last modified 12 December 2016. https://climatedataguide.ucar.edu/climate-data/atmospheric-reanalysis-overview-comparison-tables.

Gelaro, R, W. McCarty, M. J. Suárez, R. Todling, A. Molod, L. Takacs, C. A. Randles, A. Darmenov, M. G. Bosilovich, R. Reichle, K. Wargan, L. Coy, R. Cullather, C. Draper, S. Akella, V. Buchard, A. Conaty, A. M. da Silva, W. Gu, G.-K. Kim, R. Koster, R. Lucchesi, D. Merkova, J. E. Nielsen, G. Partyka, S. Pawson, W. Putman, M. Rienecker, S. D. Schubert, M. Sienkiewicz, and B. Zhao. 2017. "The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2)." *Journal of Climate* 30:5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1.

Global Modeling and Assimilation Office. 2016. *MERRA-2: File Specification*. GMAO Office Note No. 9 (Version 1.1). Greenbelt, MD: Global Modeling and Assimilation Office.

Menne, M. J., I. Durre, B. Korzeniewski, S. McNeal, K. Thomas, X. Yin, S. Anthony, R. Ray, R. S. Vose, B. E. Gleason, and T. G. Houston. 2012a. *Global Historical Climatology Network - Daily (GHCN-Daily), Version 3.24*. Accessed 28 March 2018. NOAA National Climatic Data Center. http://doi.org/10.7289/V5D21VHZ.

Menne, M. J., I. Durre, R. S. Vose, B. E. Gleason, and T. G. Houston. 2012b. "An Overview of the Global Historical Climatology Network-Daily Database." *Journal of Atmospheric and Oceanic Technology* 29:897–910

NASA (National Aeronautics and Space Administration). 2019. "Modern-Era Retrospective Analysis for Research and Applications, Version 2." Global Modeling and Assimilation Office. Last modified 1 February 2019. https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/.

Naval Facilities Engineering Command. 2016. *Pavement Design for Roads and Parking Areas*. Unified Facilities Criteria (UFC) 3-250-01. Washington, DC: U.S. Army Corps of Engineers, Naval Facilities Engineering Command, and U.S. Air Force Civil Engineer Center.

NCEI (National Centers for Environmental Information). 2010. *Federal Climate Complex Global Surface Summary of Day Data, Version 7*. ftp://ftp.ncdc.noaa.gov/pub/data/gsod/readme.txt.

NOAA (National Oceanic and Atmospheric Administration). 2018a. *Global Surface Summary of the Day (GSOD)*. Washington, DC: National Centers for Environmental Information. https://www7.ncdc.noaa.gov/CDO/cdoselect.cmd?datasetabbv=GSOD&countryabbv=&georegionabbv=.

———. 2018b. *Integrated Surface Database*. Washington, DC: National Centers for Environmental Information. https://www.ncdc.noaa.gov/isd.

———. 2018c. *Brief Review of Problems and Issues with Integrated Surface Data (ISD)*. Washington, DC: National Centers for Environmental Information. ftp://ftp.ncdc.noaa.gov/pub/data/noaa/isd-problems.pdf.

———. 2018d. "ISD References." Washington, DC: National Centers for Environmental Information. https://www.ncdc.noaa.gov/isd/references.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. 1992. *Numerical Recipes in Fortran 77: The Art of Scientific Computing*. 2nd ed. New York: Cambridge University Press.

Reinecker, M. M., M. J. Suarez, R. Gelaro, R. Todling, J. Bacmeister, E. Liu, M. G. Bosilovich, S. D. Schubert, L. Takacs, G. Kim, S. Bloom, J. Chen, D. Collins, A. Conaty, A. da Silva, W. Gu, J. Joiner, R. D. Koster, R. Lucchesi, A. Molod, T. Owens, S. Pawson, P. Pegion, C. R. Redder, R. Reichle, F. R. Robertson, A. G. Ruddick, M. Sienkiewicz, and J. Woollen. 2011. "MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications." *Journal of Climate* 24:3624–3648. https://doi.org/10.1175/JCLI-D-11-00015.1.

Schwartz, C. W., G. E. Elkins, R. Li, B. A. Visintine, B. Forman, G. R. Rada, and J. L. Groeger. 2015. *Evaluation of Long-Term Pavement Performance (LTTP) Climatic Data for Use in Mechanistic-Empirical Pavement Design Guide (MEPDG) Calibration and Other Pavement Analysis*. FHWA-HRT-15-019. McLean, VA: Federal Highway Administration, Office of Infrastructure Research and Development.

USACE (U.S. Army Corps of Engineers). 1987. *HECLIB Volume 1: HECLIB Subroutines, Programmer's Manual*. CPD-58. Davis, CA: Hydrologic Engineering Center.

———. 1991. *HECLIB Volume 2: HEC-DSS Subroutines, Programmer's Manual*. CPD-57. Davis, CA: Hydrologic Engineering Center.

———. 2001. *Pavement Design for Airfields*. UFC 3-260-02. Washington, DC: U.S. Army Corps of Engineers, Naval Facilities Engineering Command, and U.S. Air Force Civil Engineer Support Agency

———. 2004. *Calculation Methods for Determination of Depth of Freeze and Thaw in Soil: Arctic and Subarctic Construction*. UFC 3-130-06. Washington, DC: U.S. Army Corps of Engineers, Naval Facilities Engineering Command, and U.S. Air Force Civil Engineer Support Agency.

———. 2009. *HEC-DSSVue, HEC Data Storage System, Visual Utility Engine, User's Manual, Version 2.0*. CPD-79. Davis, CA: Hydrologic Engineering Center.

WorldIndex. 1999. "About WorldIndex." Help file included in WorldIndex Version 99.1. Hanover, NH: U.S. Army Cold Regions Research and Engineering Laboratory.

Appendix A: FDD Determination

In designing permanent pavements, the design freezing index (DFI), is commonly used. (DFI is equivalent to *freezing degree-day* (FDD), a common term.) The cumulative degree-days (CDD) curve is used to estimate the FDD. (From this point further, only the term FDD will be used and not DFI as FDD is consistent with the terminology of the WorldIndex database.) A degree-day is defined as "The degree-days for any one day equals the difference between the average daily air temperature and 32°F. The degree-days are negative when the average daily temperature is below 32°F (freezing degree-days) and positive when above (thawing degree-days). Degree-days may be computed in either Fahrenheit or Celsius units" (USACE 2004). The CDD curve is developed by summing the degree-days of each day starting at an appropriate time of year and proceeding day by day throughout the year. At the start of the following year, the CDD curve is reset to zero and the procedure repeated. Section 6.13 describes the process for determining the CDD curve.

The principle function of the CDD curve is to provide a good estimate of the FDD that occur during any winter. This requires that the starting point in the year for the accumulation process be before the onset of subfreezing winter temperatures; but beyond that, the actual start date is somewhat arbitrary. This update to the WorldIndex database used a start date of 1 August for the northern hemisphere and 1 February for the southern hemisphere.

The FDD for a winter are equal to the difference between the maximum CDD, CDD_{max}, and minimum CDD, CDD_{min}, which are associated with the subfreezing temperatures of the winter season (see Figure A-1). CDD_{max} and CDD_{min} are inflection points in the CDD time series. Note that CDD_{max} is not necessarily the maximum CDD that occurs over the 1-year period starting on 1 August (for the northern hemisphere). Rather, CDD_{max} is the maximum CDD that occurs at the start of the period of subfreezing temperatures associated with the winter season. In a similar way, CDD_{min} is not necessarily the minimum CDD that occurs over the 1-year period starting on 1 August (for the northern hemisphere). Rather, CDD_{min} is the minimum CDD that occurs at the end of the period of subfreezing temperatures associated with the winter season. As a result, the accurate identification of CDD_{max} and CDD_{min} requires that the period of subfreezing temperatures be determined first.

The following procedure is used to determine the FDD for each winter season. This procedure is designed to provide a robust estimator of the inflection points over a wide range of climatic conditions. It fits a sine curve with a 1-year period to the time series of the daily average temperature. This allows the end of the winter season to be determined with some precision for each year. The maximum inflection point is then found by searching **CDD** (*j*) from the end of the winter back to the start of the data series on 1 August (for the northern hemisphere). The minimum inflection point is then found by searching **CDD** (*j*) from the date of the maximum inflection point to the end of the data series on 31 July (for the northern hemisphere).

The data for a winter season for any given year begins on 1 August and extends to 31 July of the next year (for the northern hemisphere). The average daily temperature has been estimated for each day prior to the start of this procedure. If a day had missing data, the average daily temperature based on climatology was used as described in section 6.1.2.

Next, a sine curve is fitted to the time series of daily average temperatures. First, the average daily temperature over this year, **yearly_avg** (*year*), is subtracted from the daily average air temperature for each day to produce a new time series with zero mean, Y_{obs} (*j*):

$$Y_{obs}\left(j\right) = \mathbf{T_{avg}}(j) - yearly\_avg\left(year\right).$$

The variance of Y_{obs} (*j*) is found as

$$V_{Y_{obs}} = \frac{1}{\left(n-1\right)} \sum_{j=1}^{n} Y_{obs}^{2}\left(j\right).$$

It is assumed that Y_{obs} (*j*) can be represented by a sine curve with a period of 1 year,

$$Y^{*}(j) = a_1 \sin\left(\frac{2\pi j}{n} - \alpha_1\right),$$

where

n = the number of days in a year,
a_1 = the amplitude of the sine curve, and
α_1 = its phase angle.

The variance of $Y^*(j)$ is found as

$$V_{Y^*} = \frac{1}{(n-1)} \sum_{j=1}^{n} Y^{*2}(j) = \frac{1}{(n-1)} \int_1^n \left(a_1 \sin\left(\frac{2\pi x}{n} \right) \right)^2 dx = \frac{n}{(n-1)} \frac{a_1^2}{2}.$$

Note that in estimating the variance of the sine curve, the phase angle was assumed equal to zero with no impact on the results. The amplitude, a_1, of the fitted sine curve can then be estimated by equating the variances as

$$\hat{a}_1 = \sqrt{2V_{Y_{obs}} \frac{n-1}{n}},$$

where $\hat{a}_1$ is the estimated amplitude. Next the estimated value of the phase angle is estimated by finding the phase angle that minimizes, $M(j_o)$, found as

$$M(j_o) = \sum_{j=1}^{n} \left(Y_{obs}(j) - Y^*(j) \right)^2 = \sum_{j=1}^{n} \left(Y_{obs}(j) - \hat{a}_1 \sin\left(\frac{2\pi j}{n} - \hat{\alpha} \right) \right)^2,$$

where

$$\hat{\alpha} = \frac{2\pi j_o}{n}$$

and $1 \le j_o \le n$. The search is conducted by iterating j_o from 1 to n in steps of 1 day to find the value of j_o that produces the minimum of M. At the end of this procedure, the amplitude, $\hat{a}_1$, the phase angle, $\hat{\alpha}$, and the average, T_{avgFDD} are all known. The end of the period of subfreezing temperatures associated with the winter season can then be estimated as

$$j_{base} = \frac{n}{2\pi} \left(\hat{\alpha} + \sin^{-1}\left(\frac{T_{base} - T_{avgFDD}}{\hat{a}_1} \right) \right),$$

where j_{base} = the day of the year (1 August = 1) that corresponds to the end of the period of subfreezing temperatures associated with the winter season and T_{avgFDD} = *yearly_avg* (*year*), the average annual temperature.

Note that $\sin^{-1}$ is defined only if

$$-1 \le \left(\frac{T_{base} - T_{avgFDD}}{\hat{a}_1} \right) \le 1 .$$

If this is not true, then j_{base} is returned as the last day of the year with $\mathbf{T_{avg}}$ less than or equal to T_{base}.

CDD_{max} is found in the series **CDD** (1) to **CDD** (j_{base}) by iterating from j_{base} to 1 in steps of −1 day to find the maximum value of CDD for the year. The value of CDD_{max} is then stored in the array $\mathbf{CDD_{max}}$ (year). The Julian date on which CDD_{max} occurs is recorded in the array, $\mathbf{j_{max}}$ (*year*).

CDD_{min} is found in the series **CDD** ($\mathbf{j_{max}}$ (*year*)) to **CDD** (*n*) by iterating *j* from $\mathbf{j_{max}}$ (*year*) to *n* in steps of 1 day to find the minimum value of CDD for the year. The value of CDD_{min} is then stored in the array $\mathbf{CDD_{min}}$ (year). The Julian date CDD_{min} occurs on is also recorded, $\mathbf{j_{min}}$ (*year*).

The next two figures show example calculations. Figure A-1 shows the recorded daily average air temperature for 1 year at a specific location (Minot, ND). The blue line is the fitted sine curve. The black line is the computed CDD. The estimate of the end of the period of subfreezing temperatures associated with the winter season is indicated by the red dot. The red dot is located on the sine curve at 0°C. Note that there are two locations where the sine curve passes through 0°C. Only the time where the sine curve is increasing in value as time increases is selected. This time is not used directly; rather, it used as the point in the year to begin the search for CDD_{max}. The search begins at the red dot and proceeds backwards in time. CDD_{max} is the maximum CDD value between the start of the year and the time of the red dot, which indicates the start of the freezing season. The date of CDD_{max} is $\mathbf{j_{max}}$ (*year*). The search for CDD_{min} starts at $\mathbf{j_{max}}$ (*year*) and proceeds forward in time. CDD_{min} is the minimum CDD value between $\mathbf{j_{max}}$ (*year*) and the end of the year and indicates the end of the freezing season. The date of CDD_{min} is $\mathbf{j_{min}}$ (*year*).

$\mathbf{j_{max}}$ (*year*) and $\mathbf{j_{min}}$ (*year*) are used in determining several other temperature parameters. $\mathbf{j_{max}}$ (*year*) is used in defining the start of the freezing season in the following WorldIndex parameters: AVERAGE START, STDEV START, LATE START, and EARLY START. $\mathbf{j_{min}}$ (*year*) is used in defining the end of the freezing season in the following WorldIndex parameters: AVERAGE STOP, STDEV STOP, LATE STOP, and EARLY

STOP. The difference between $\mathbf{j_{max}}$ (*year*) and $\mathbf{j_{min}}$ (*year*) is used in defining the length of the freezing season in MLFS and STDEV LFS.

Figure A-1. Recorded daily average air temperature, fitted sine curve, and CDD for a 1-year period in Minot, ND. The *red dot* indicates the estimate for the end of the period of subfreezing temperatures associated with the winter season.

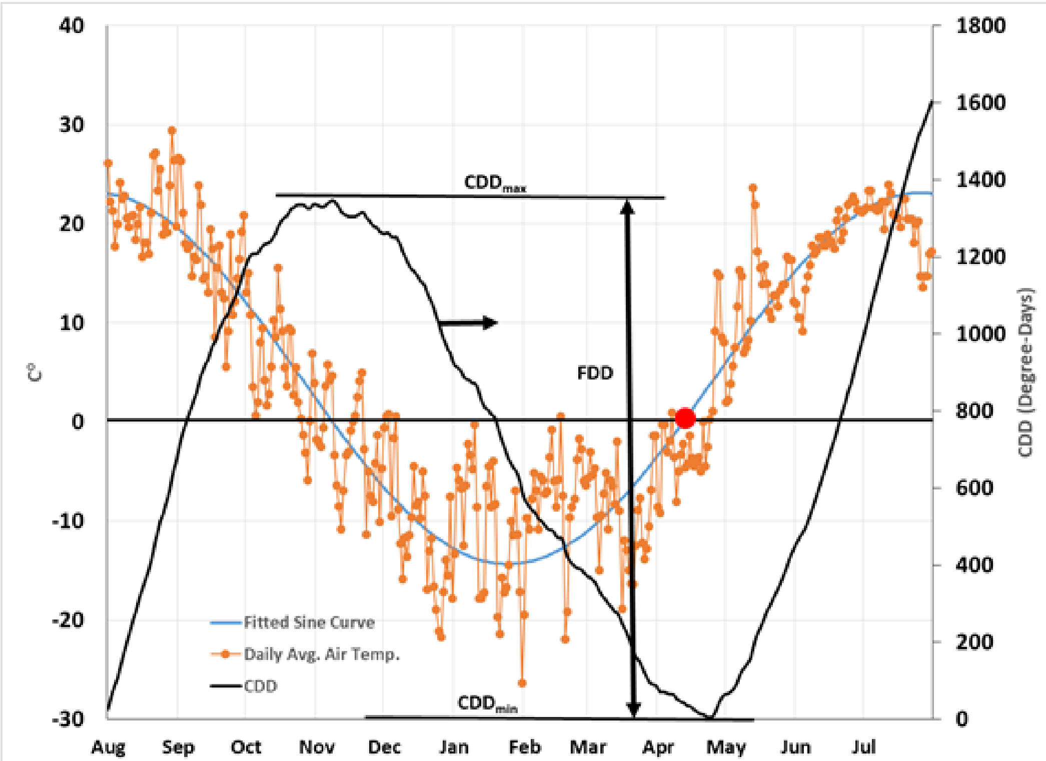
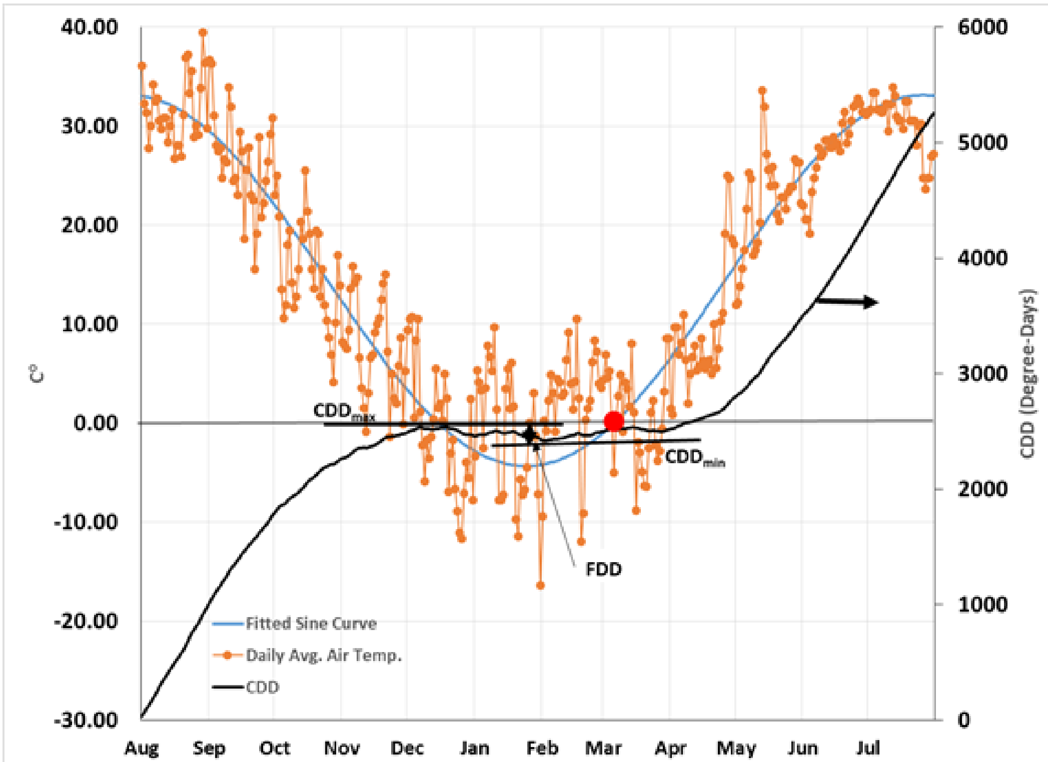


Figure A-2 shows the daily average air temperature for 1 year at a relatively warm location. (The air-temperature data was created especially for this example.) As before, the fitted sine curve and computed CDD are also shown. The end of the winter season would be relatively difficult to determine by examining the daily air-temperature data alone due to the high variability of the daily average air temperature during the winter season. The fitted sine curve provides a well-defined location to begin the search for CDD_{max} and then CDD_{min}.

Figure A-2. Recorded daily average air temperature, fitted sine curve, and CDD for a 1-year period for a relatively warm location. The *red dot* indicates the estimate for the end of the period of subfreezing temperatures associated with the winter season.



REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
July 2019

2. REPORT TYPE
Technical Report/Final

3. DATES COVERED *(From - To)*

4. TITLE AND SUBTITLE
WorldIndex Database Update 2018

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Lynette A. Barna, Steven F. Daly, Kathleen F. Jones, and John J. Gagnon

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Engineer Research and Development Center (ERDC)
Cold Regions Research and Engineering Laboratory (CRREL)
72 Lyme Road
Hanover, NH 03755-1290

8. PERFORMING ORGANIZATION REPORT NUMBER
ERDC/CRREL TR-19-13

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Corps of Engineers, Transportation Systems Center
1616 Capitol Avenue
Omaha, NE 68102-4901

10. SPONSOR/MONITOR'S ACRONYM(S)
USACE

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
Pavement-Transportation Computer Assisted Structural Engineering (PCASE). Funding provided by FAN S0140441.

14. ABSTRACT

The primary software program for designing and evaluating pavements on military installations is the Pavement-Transportation Computer Assisted Structural Engineering (PCASE) program. PCASE, which is undergoing an update to version PCASE7, is an engineering tool for the design and evaluation of airfields and roadways using Department of Defense criteria. The WorldIndex database is a necessary component of PCASE and is integrated into it. The WorldIndex database internally supplies PCASE with the climate parameters required to determine the maximum depth of frost penetration. The database contains other temperature-based parameters used in structural evaluations.

This report documents a recent update to the WorldIndex database that uses historical surface air-temperature observations from 1980 to 2017 at over 10,000 l0cations around the globe. The database contains 80 air-temperature-based parameters determined for each station, including the three parameters that PCASE requires to compute the maximum depth of frost penetration for pavement design in cold regions: the average annual air temperature, the average freezing degree-days, and the mean length of the freezing season. The report concludes with recommendations for future updates.

15. SUBJECT TERMS
Atmospheric temperature, Climate, Degree days, Military bases, Pavements, Pavements--Effect of temperature on, Pavements--Frost damage, Pavement-Transportation Computer Assisted Structural Engineering (PCASE), Thawing, WorldIndex database

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON / 19b. TELEPHONE NUMBER *(include area code)*
Unclassified	Unclassified	Unclassified	SAR	55	